

The span of correlations in dolphin whistle sequences

Ramon Ferrer-i-Cancho^{1,*} and Brenda McCowan²

¹ Complexity & Quantitative Linguistics Lab
 Departament de Llenguatges i Sistemes Informàtics,
 TALP Research Center, Universitat Politècnica de Catalunya,
 Campus Nord, Edifici Omega Jordi Girona Salgado 1-3.
 08034 Barcelona, Catalonia (Spain)

² Population Health & Reproduction,
 School of Veterinary Medicine,
 University of California,
 One Shields Ave 1024 Haring Hall
 Davis, CA 95616, United States

E-mail: rferrericancho@lsi.upc.edu and bjmcowan@ucdavis.edu

Abstract. Long-range correlations are found in symbolic sequences from human language, music and DNA. Determining the span of correlations in dolphin whistle sequences is crucial for shedding light on their communicative complexity. Dolphin whistles share various statistical properties with human words, i.e. Zipf's law for word frequencies (namely that the probability of the i th most frequent word of a text is about $i^{-\alpha}$) and a parallel of the tendency of more frequent words to have more meanings. The finding of Zipf's law for word frequencies in dolphin whistles has been the topic of an intense debate on its implications. One of the major arguments against the relevance of Zipf's law in dolphin whistles is that is not possible to distinguish the outcome of a die rolling experiment from that of a linguistic or communicative source producing Zipf's law for word frequencies. Here we show that statistically significant whistle-whistle correlations extend back to the 2nd previous whistle in the sequence using a global randomization test and to the 4th previous whistle using a local randomization test. None of these correlations are expected by a die rolling experiment and other simple explanation of Zipf's law for word frequencies such as Simon's model that produce sequences of unpredictable elements.

Keywords: Zipf's law, die rolling, random typing, dolphin whistles.

PACS numbers: 89.70.-a Information and communication theory
 89.75.Da Systems obeying scaling laws
 05.45.Tp Time series analysis

1. Introduction

Long-range correlations have been reported in different kinds of symbolic sequences: human language [1, 2, 3, 4], DNA [5] and music [6, 4]. A few studies have studied the span of correlations in sequences of behavior produced by other species [7, 8, 9]. Rather long-correlations (extending back at least to the 7th previous element) have reported in sequences of dolphin surface behavioral patterns [7]. A preliminary analysis of constraints in sequences of dolphin whistles was performed in Ref. [8] but strong conclusions were not reached due to the small size of the dataset. Determining the actual span of correlations in dolphin sequences is crucial for shedding light on the communicative complexity of dolphins whistles.

Various similarities between human words and dolphin whistles have been reported. A parallel of Zipf's law of meaning distribution, the tendency or more frequent words to have more meanings [10], has been found in dolphins whistles [11]. Zipf's law for word frequencies, namely, the probability of the i th most frequent word of a text, is $\sim i^{-\alpha}$, where α is the exponent of the law [10], has also been found in dolphin whistles [8]. However, the finding remains controversial because it has been argued that simply rolling a die could explain the presence of the law in dolphin whistles [12]. The experiment consists of generating a sequence of faces by rolling a die. One of the faces of the die plays the role of a word delimiter. For instance, a die of six faces could produce 1, 5, 2, 6, 3, 2, 4, 3, 5, 6... Treating 6 as a pseudo-word delimiter, the previous sequence of faces becomes the sequence of pseudo-words 152, 32435, The experiment is an abstraction of the popular monkey typing experiment, that consists of typing at random on a keyboard (the face that plays the role of the word delimiter is the space and the other faces are letters) [13]. Die rolling and monkey typing have been argued to explain or mimic Zipf's law for word frequencies in human words [13] and dolphin whistles [12]. Another die rolling experiment, where the probability of the i th face is the probability of the i th most frequent word, has been proposed [14].

The hypothesis of die rolling as an explanation for Zipf's law for word frequencies in human language in other species can be tested in at least two different ways:

- (i) *By comparing the actual distribution of 'word' frequencies with the one that is actually produced by the die rolling process [15].* Concerning the die rolling experiment of Ref. [12], the parameters of the model that provide a satisfactory fit to actual word frequencies according to a statistically rigorous test, are indeed unknown [15], in spite of the many previous claims about the good fit of the model using qualitative arguments [13, 12]. Concerning the die rolling experiment or Ref. [14], the model is able to trivially provide a perfect fit to any theoretical or empirical discrete distribution, being Zipf's law for word frequencies a particular case.
- (ii) *By comparing the statistical properties of the sequence of words produced with those of the sequence that is produced by the die rolling process.* The rolling experiments of Refs. [12, 14] produce a sequence of independent 'words' in the sense that elements that have already been produced carry no information about the next element. In

contrast human language, shows long range correlations in texts using words (e.g. [2, 3]) or letters (e.g., [1, 16]) as units of the sequence.

Here we will follow the second track for dolphin whistles. The aim of the present article is to determine the span of correlations in dolphin whistle sequences and evaluate the suitability of die rolling [12, 14]. The challenge of the analysis is facing the statistical problems arising from the rather small size of the dataset of Ref. [8]. The danger of undersampling in the context of dolphins whistles has already been discussed [8]. The next section presents the information theoretic measure approach that will be used to study correlations in dolphin whistle sequences in that dataset.

2. Mutual information

We consider pairs of whistles in a sequence and three related random variables: X for the whistle that is the 1st member of the pair, Y for the whistle that is the second member of the pair and D for their distance. We adopt the convention that consecutive elements are at distance 1, elements separated by one element are at distance 2, and so on...[17]. Given a collection of sequences of whistles, $p(X = x, Y = y|D = d)$ is defined as the probability that whistle x is followed by whistle y knowing that they are at distance d from each other and d_{max} is defined as the maximum distance considered in the analysis (thus $1 \leq d \leq d_{max}$). Given a certain distance d , the marginal conditional probabilities are defined as

$$p(X = x|D = d) = \sum_y p(X = x, Y = y|D = d)$$

$$p(Y = y|D = d) = \sum_x p(X = x, Y = y|D = d).$$

$I(X; Y|D = d)$, the conditional mutual information between X and Y given a concrete distance d , is defined as [18]

$$I(X; Y|D = d) = \sum_{x,y} p(X = x, Y = y|D = d) \log q(X = x, Y = y|D = d), \quad (1)$$

where

$$q(X = x, Y = y|D = d) = \frac{p(X = x, Y = y|D = d)}{p(X = x|D = d)p(Y = y|D = d)}.$$

$I(X; Y|D = d)$ has been used to study long-range correlations in DNA, texts and music [5, 4].

$f(X = x, Y = y|D = d)$ is defined as the number of times that x has been followed by y at distance d . The marginal conditional frequencies are defined as

$$f(X = x|D = d) = \sum_y f(X = x, Y = y|D = d) \quad (2)$$

$$f(Y = y|D = d) = \sum_x f(X = x, Y = y|D = d), \quad (3)$$

and the total number of pairs at distance d is defined as

$$F(d) = \sum_x f(X = x|D = d) = \sum_y f(Y = y|D = d).$$

In a finite collection of sequences, $p(X = x, Y = y|D = d)$, $p(X = x|D = d)$ and $p(Y = y|D = d)$ are relative frequencies, i.e.

$$p(X = x, Y = y|D = d) = \frac{f(X = x, Y = y|D = d)}{F(d)} \quad (4)$$

$$p(X = x|D = d) = \frac{f(X = x|D = d)}{F(d)} \quad (5)$$

$$p(Y = y|D = d) = \frac{f(Y = y|D = d)}{F(d)}. \quad (6)$$

In a real collection of sequences, $p(X = x, Y = y|D = d)$, $p(X = x|D = d)$ and $p(Y = y|D = d)$ are estimated from these relative frequencies. Applying Eqs. 4, 5 and 6 to Eq. 1, yields the *sample* mutual information between X and Y given a concrete distance d ,

$$I_s(X; Y|D = d) = \log F(d) + \frac{\sigma_1 - \sigma_2 - \sigma_3}{F(d)}, \quad (7)$$

where

$$\begin{aligned} \sigma_1 &= \sum_{x,y} f(X = x, Y = y|D = d) \log f(X = x, Y = y|D = d) \\ \sigma_2 &= \sum_x f(X = x|D = d) \log f(X = x|D = d) \\ \sigma_3 &= \sum_y f(Y = y|D = d) \log f(Y = y|D = d). \end{aligned}$$

Next two useful properties of $I_s(X; Y|D = d)$ are presented:

- (i) If marginal conditional frequencies at distance d are boolean, i.e. $f(X = x|D = d) = f(Y = y|D = d) \in \{0, 1\}$ for any x and y then $I_s(X; Y|D = d) = \log F(d)$. To see it, notice that $f(X = x|D = d) = f(Y = y|D = d) = 1$ implies, thanks to Eqs. 2 and 3, that the joint frequencies are all boolean, i.e. $f(X = x, Y = y|D = d) \in \{0, 1\}$ for any x and y . According to Eq. 7, the fact that both marginal and joint conditional frequencies do not exceed one yields $\sigma_1 = \sigma_2 = \sigma_3 = 0$ and thus $I_s(X; Y|D = d) = \log F(d)$ as we wanted to prove.
- (ii) If all whistles at cooccurring at distance d are identical (only one whistle type has non-zero frequency), then $\sigma_1 = \sigma_2 = \sigma_3 = F(d) \log F(d)$ and then, according to Eq. 7, $I_s(X; Y|D = d) = 0$.

3. Methods

We reused the collection of whistle sequences employed to study Zipf's law in dolphin whistles [8]. A summary of the elementary statistical properties of the sequences is provided in Table 1. Sequences of length smaller than 2 were filtered out.

3.1. Distances where correlations are significant

For each dolphin in the dataset, his/her collection of sequences was analyzed to extract a list of distances in the interval $[2, d_{max}^2]$ at which $I_s(X; Y|D = d)$ is statistically

significant at a significance level of $\alpha = 0.05$. For each dolphin and each distance, we used a Monte Carlo procedure for estimating a p -value indicating the probability that the value of $I_s(X; Y|D = d)$ from a randomized version of the data $I_s(X; Y|D = d)$ is at least as large as that of the original collection of sequences:

- (i) $R = 10^7$ randomized versions of the original data were generated.
- (ii) R_{\geq} , the number of times the value of $I_s(X; Y|D = d)$ is at least as large as that of the original data was calculated for each $d \in [1, d_{max}^2]$ and the p -value was estimated as R_{\geq}/R .
- (iii) For each dolphin, all distances d such that $p\text{-value} = R_{\geq}/R \leq \alpha$ were added to the list of distances.

3.2. Upper bounds for the span of correlations

We say that $I_s(X; Y|D = d)$ is constant if $I_s(X; Y|D = d)$ is the same for any randomization of data upon which $I_s(X; Y|D = d)$ is computed. We consider two situations in which $I_s(X; Y|D = d)$ cannot be significantly high: (a) $I_s(X; Y|D = d)$ is constant in the sense above or (b) $I_s(X; Y; D = d) = 0$ as 0 is actually the minimum of mutual information [18]. Accordingly, we consider three different ways of defining d_{max} :

- d_{max}^0 , defined through the number of pairs of elements at distance d in a sequence of length l , i.e. $\pi(d, l) = l - d$. The maximum distance d that can be considered in a collection of sequences where the longest sequence has length l_{max} is obtained from $\pi(d_{max}, l_{max}) \geq 1$, which gives $d_{max}^0 = l_{max} - 1$.
- d_{max}^1 , defined as the smallest distance at which $I_s(X; Y|D = d)$ is constant in the sense above for any $d \in (d_{max}^1, d_{max}^0]$.
- d_{max}^2 , defined as the smallest distance beyond which $I_s(X; Y|D = d)$ cannot be significantly high for $d \in (d_{max}^2, d_{max}^0]$. Notice that $d_{max}^2 \leq d_{max}^1$ because $I_s(X; Y|D = d)$ cannot be significantly high if $I_s(X; Y|D = d)$ is constant.

The interest of these definitions of d_{max} is two-fold. First, bounding a priori the span of correlations between whistles. Second, reducing the computational cost of evaluating the significance of $I_s(X; Y|D = d)$ at distances where the result of the test is straightforward. Distances greater than d_{max}^2 can be discarded.

We did not check if $I_s(X; Y|D = d) = 0$ using Eq. 7 as this is problematic due to finite numerical precision for real numbers. Eq. 1 indicates that $I_s(X; Y|D = d) = 0$ if and only if $q(X = x; Y = y|D = d) = 1$ for any x and y such that $p(X = x; Y = y|D = d) > 0$. Applying Eqs. 4, 5 and 6, it is easy to see that the condition $q(X = x; Y = y|D = d) = 1$ is equivalent to a more numerically convenient condition, i.e. $F(d)f(X = x, Y = y|D = d) = f(X = x|D = d)f(Y = y|D = d)$.

3.3. Kinds of randomization

Two kind of randomizations of the data upon $I_s(X; Y|D = d)$ is calculated were considered: global and local randomization. A precise calculation of d_{max}^1 and d_{max}^2

Table 1. Summary of the elementary statistical properties of the collections of sequences of each dolphin. For each dolphin, the following information is shown: dolphin's name, T , the total number of whistle types, V , the number of different whistle types, S (the number of sequences), $\langle l \rangle = T/S$ (the mean sequence length in whistle types), l_{max} (the maximum length). Dolphins are sorted decreasingly by T .

Name	T	V	S	$\langle l \rangle$	l_{max}
Liberty	334	31	102	3.27	11
Norman	235	39	74	3.18	21
Panama	148	42	43	3.44	10
Chelsea	110	12	34	3.24	8
Sadie	110	13	33	3.33	8
Stormy	94	12	27	3.48	9
Delphi	65	18	24	2.71	7
Neptune	53	9	9	5.89	13
Circe	47	4	15	3.13	7
Desmond	40	3	7	5.71	10
Sam	39	4	12	3.25	13
Tasha	35	9	12	2.92	7
Bayou	27	10	8	3.38	8
Terry	21	7	5	4.20	7
Schooner	19	4	6	3.17	6
ECB	15	5	6	2.50	4
Gordo	6	6	2	3.00	4

depends on the kind of randomization.

3.3.1. Global randomization Here the data that is randomized is the whole collection of sequences. The randomization procedure consisted of copying all whistles in a vector of length

$$T = \sum_{i=1}^S l_i,$$

where l_i is the length of the i th sequence and S is the number of sequences of the collection. Then, every randomized version of the collection is obtained by generating a uniformly distributed random permutation of the vector [19] and cutting that vector in pieces of lengths $l_1, \dots, l_i, \dots, l_S$ (cut always in this order) to produce the randomized collection of sequences. Notice that the randomization procedure preserves the sequence lengths and the frequencies of each whistle in the original collection of sequences. Concerning the computation of d_{max}^1 , notice that constant $I_s(X; Y|D = d)$ with regard to any global randomization can occur in three circumstances:

- (i) If all the whistle types are hapax legomena, i.e. they occur only once in the whole collection as all the marginal conditional frequencies are boolean and thus $I(X; Y|D = d) = \log F(d)$, as it has been shown above, for any $d \in [1, d_{max}^0]$.

- (ii) If the repertoire size is one (there is only one whistle type of non-zero frequency) then all the whistles cooccurring at a certain distance are identical and thus $I(X; Y|D = d) = 0$, as it has been shown above, for any $d \in [1, d_{max}^0]$.
- (iii) When $F(d) = 1$, all the marginal conditional frequencies are boolean in this case and then $I(X; Y|D = d) = \log F(d) = \log 1 = 0$.

These conditions lead to the following specific procedure for calculating d_{max}^1 . Satisfying condition (i) (this is the case of the dolphin 'Gordo') or (ii) implies $d_{max}^1 = 0$. If none of these two conditions is met, d_{max}^1 is calculated by means of condition (iii). In order to warrant $F(d) > 1$, d_{max}^1 must be d_{max}^0 if there is more than one sequence of maximum length and $d_{max}^1 = d_{max}^0 - 1$ otherwise. To see it, notice that if the collection has S_{max} sequences of maximum length, then $F(d_{max}^1) = S_{max}\pi(l_{max} - 1, l) = S_{max}$ and thus $F(d) \geq 2$ if and only if $S_{max} \geq 2$.

3.3.2. Local randomization Here the data that is randomized are the pairs of whistles occurring at a certain distance d . The randomization procedure consisted of copying all the pairs of whistles occurring at distance d in a vector of length $2F(d)$. Then, every randomized version of the collection was obtained by generating a uniformly distributed random permutation of the vector [19] and then taking the i th and the $(i+1)$ th whistles of the vector, with $1 \leq i \leq 2F(d)$, to form the j th pair of whistles of the randomized pairs of whistles, with $1 \leq j \leq F(d)$ and $j = \lceil i/2 \rceil$. Notice that the randomization procedure preserves the frequencies of each whistle in the original pairs of whistles at distance d .

Concerning the computation of d_{max}^1 for local normalization notice that constant $I_s(X; Y|D = d)$ with regard to any local randomization for distance d , adds two new relevant conditions with regard to global randomization:

- (iv) When all whistle types are locally hapax legomena, i.e. all the whistle types forming the pairs at distance d occur only once in the ensemble of pairs at distance d . In that case, all marginal conditional frequencies are boolean and thus $I_s(X; Y; D = d) = F(d)$ as it has been shown above.
- (v) When the local repertoire has size one, i.e. all the pairs at distance d are made of a single whistle type. In that case, all $I_s(X; Y; D = d) = 0$ as it has been shown above.

The procedure for calculating d_{max}^1 under local normalization is the following. First, compute d_{max}^1 with the same procedure used for global normalization. Decrease d_{max}^1 while $d_{max}^1 > 0$ and at least one of the two following conditions is met at distance d_{max}^1 : (a) all whistle types are locally hapax legomena or (b) the local repertoire size is one.

4. Results

Table 2 shows the values of d where $I_s(X; Y|D = d)$ is significantly high in the dolphins from Ref. [8] at a significance level of 0.05 for global and local randomization. n_d^a is

Table 2. Analysis of the span of correlation between whistles at a certain distance d using global and local randomization. For the collection of whistles sequences of each dolphin, the following information is shown: the dolphin's name, $d_{max}^{0,1,2}$ (the maximum distance according to different definitions) and the list of distances $d \in [1, d_{max}^2]$ such that $I_s(X; Y|D = d)$ is significantly high at a significance level of 0.05. Dolphins are sorted decreasingly by T .

Name	Global randomization				Local randomization		
	d_{max}^0	d_{max}^1	d_{max}^2	Distances	d_{max}^1	d_{max}^2	Distances
Liberty	10	9	7		8	7	1, 4, 6, 7
Norman	20	19	15	1, 2	19	15	1, 2, 3, 4
Panama	9	8	7	5	8	7	1, 2, 3
Chelsea	7	6	3	1	4	3	1, 2
Sadie	7	6	6	1, 2	6	6	1, 2
Stormy	8	7	6	1	7	6	1, 2, 3
Delphi	6	5	5		4	4	1
Neptune	12	11	7	2	11	7	1, 2, 3, 4
Circe	6	5	1		1	1	
Desmond	9	8	3	1	8	3	
Sam	12	11	9	7	10	9	
Tasha	6	5	5		5	5	
Bayou	7	6	6		5	5	1, 2
Terry	6	5	3	1	4	3	1
Schooner	5	4	2	1	4	2	1
ECB	3	2	2		2	2	
Gordo	3	0	0		0	0	

defined as the number of dolphins for which $I_s(X; Y|D = d)$ is significantly high at a significance level a . Table 2 yields $n_1^{0.05} = 7$, $n_2^{0.05} = 3$, $n_5^{0.05} = n_7^{0.05} = 1$ for global randomization and $n_1^{0.05} = 11$, $n_2^{0.05} = 7$, $n_3^{0.05} = 4$, $n_4^{0.05} = 3$, $n_6^{0.05} = n_7^{0.05} = 1$ for local randomization.

Next we will perform a meta-analysis to determine when $n_d^{0.05}$ is significantly high because $n_d^{0.05}$ could be the outcome of false positives (type I statistical errors) from the test that determines if $I_s(X; Y|D = d)$ is significantly high for a certain dolphin. The p -value of a continuous statistic is known to be uniformly distributed under the null hypothesis [20]. In our case, $I_s(X; Y|D = d)$ is approximately continuous and the quality of the approximation is expected to increase with the size of the permutation space of the collection of sequences. This implies that n_d^a , the number of dolphins for which $I_s(X; Y|D = d)$ is significantly high at a significance level a follows approximately a binomial distribution with parameters N_d and a , where N_d is the total number of dolphins for whom $I_s(X; Y|D = d)$ can be significantly high according to the criteria above, and $a = 0.05$ is the significance level. Notice that the dolphin 'Gordo' is a special case because he does not have distances at which mutual information can be significantly high (he has $d_{max}^1 = d_{max}^2 = 0$ because all the whistles in his collection are hapax legomena; notice $T = V$ in Table 1). Thus, N_d is not simply the number of

Table 3. Summary of the test of the significance of $n_d^{0.05}$, the number of dolphins for which $I_s(X; Y|D = d)$ is significantly high at a significance level of 0.05. Distances where $n_d^{0.05}$ cannot be significantly high (as $n_d^{0.05} = 0$ is minimum) are excluded from the analyses and filled with $-$. N_d is the number of dolphins for whom $I_s(X; Y|D = d)$ can be significantly high. The p -values indicate the probability of reaching a value of $n_d^{0.05}$ at least as high as the actual one simply by chance. The p -values were estimated by means of two different procedures: a binomial approximation and a Monte Carlo method. p -values were rounded to leave only one significant digit.

d	Global randomization				Local randomization			
	$n_d^{0.05}$	N_d	p -value		$n_d^{0.05}$	N_d	p -value	
			Binomial	Numerical			Binomial	Numerical
1	7	16	6×10^{-6}	< 0.001	11	16	2×10^{-11}	< 0.001
2	3	15	0.04	0.02	7	15	3×10^{-6}	< 0.001
3	0	—	—	—	4	13	0.003	0.001
4	0	—	—	—	3	10	0.01	0.009
5	1	10	0.4	0.3	0	—	—	—
6	0	—	—	—	1	7	0.3	0.2
7	1	5	0.2	0.2	1	5	0.2	0.2

dolphins in the dataset.

A binomial test can be used to assess if n_d^a is significantly large. The p -value of this test is

$$p - \text{value} = \sum_{x=n_d^a}^{N_d} \binom{N_d}{x} a^x (1-a)^{N_d-x}. \quad (8)$$

Concerning global randomization, a binomial test indicates that the number of dolphins showing a significant correlation is significantly high for $d = 1$ and $d = 2$ but not for $d > 2$ (Table 3). Concerning local randomization, a binomial test indicates that the number of dolphins showing a significant correlation is significantly high for $1 \leq d \leq 4$ but not for $d > 4$ (Table 3).

The meta-analysis based upon a binomial test assumes that 0.05 is the probability of rejecting the null hypothesis by chance but indeed the accuracy of the assumption depends on the properties of the collection of whistle types. For instance, we have seen that the mutual information cannot be significantly high for the dolphin 'Gordo' and thus the probability of rejecting the null hypothesis by chance is thus 0 for him. A Monte Carlo meta-analysis allows one to improve the approximate calculation of the p -value offered by Eq. 8 to some degree of numerical precision. Consider that a randomized ensemble of collections of sequences consists of a randomized version of the collection of each of the dolphins (excluding 'Gordo'). For each distance d , the p -value of n_d^a is estimated by the proportion of times in R_m randomized ensembles of collections that the value of n_d^a is equal or greater than the value of n_d^a of the original (non-randomized) ensemble. $R_m = 1000$ and $R = 1000$ were used. Concerning global randomization,

Table 4. Summary of the support for significance of each distance (based upon Table 3). $n_d^{0.05}$ is defined as the number of dolphins for which $I_s(X; Y|D = d)$ is significantly high at a significance level of 0.05. For each kind of randomization (global or local) and meta-analysis (binomial approximation or Monte Carlo method), the distances at which $n_d^{0.05}$ is significantly high are indicated.

	Distances	
	Global randomization	Local randomization
$n_d^{0.05} > 0$	1, 2, 5, 7	1, 2, 3, 4, 6, 7
Meta-analysis (binomial approx.)	1, 2	1, 2, 3, 4
Meta-analysis (Monte Carlo method)	1, 2	1, 2, 3, 4

one obtains that $n_d^{0.05}$ is significantly high for $d = 1$ and $d = 2$ but not for $d > 2$ (Table 3). Concerning local randomization, one obtains that $n_d^{0.05}$ is significantly high for $1 \leq d \leq 4$ but not for $d > 4$ (Table 3). To sum up, the conclusions of the two meta-analyses coincide from a qualitative point of view (Table 4). The binomial test provides a good enough approximation for both global and local randomization.

From a global randomization perspective, correlations are short range: the presence of significant correlations at low distances ($d = 2$ and specially $d = 1$) is unquestionable but significant correlations at higher distances could be false positives (Table 4). From a local randomization perspective, correlations extending back to the 4th back whistle type are unquestionable but farther significant correlations could be false positives (Table 4).

5. Discussion

We have demonstrated that, for the majority of individuals, a dolphin whistle carries (on average) a significant amount of information about the next whistles of the sequence (Tables 2 and 3). Global randomization indicates that a whistle carries information about at least one of the next two whistles of the sequence whereas local randomization indicates that a whistle carries information about at least one of the next four whistles of the sequence (Table 4). This is a property that is inconsistent with die-rolling, where a pseudo-word carries no information at all about the next pseudo-words of the sequence. The fact that this is also a feature also shared by Simon’s model for word frequencies [21], questions the validity of a popular argument against the utility of Zipf’s law for frequencies, namely that the law is of practically no help in assessing the complexity of a communication system because there are many ways of reproducing it [22, 23, 14, 12]. Indeed, when the statistical properties of the sequence are taken into account, the number of candidate explanations drops down. The multiplicity of explanations for Zipf’s law in a species such as dolphins depends on how many statistical features, besides Zipf’s law for word frequencies, are used to break the tie between candidates. The big question that future research on dolphins whistles must address is: what is the communicative complexity of a system whose units (e.g., whistles types) are distributed

following Zipf's law for word frequencies [8], show a parallel of Zipf's law of meaning distribution [11] and form sequences with correlations that defy a simple explanation such as die rolling or Simon's model? We hope that our research stimulates further data collection to determine if the rather short range correlation discovered here are an intrinsic property of dolphin whistle communication or a consequence of the small size of our dataset.

Acknowledgments

We are grateful to L. Doyle for helpful discussions and to M. Piattelli-Palmarini for making us aware of Ref. [14]. This work was supported by the grant *Iniciació i reincorporació a la recerca* from the Universitat Politècnica de Catalunya and the grants BASMATI (TIN2011-27479-C04-03) and OpenMT-2 (TIN2009-14675-C03) from the Spanish Ministry of Science and Innovation.

References

- [1] F. Moscoso del Prado Martín. The universal 'shape' of human languages: spectral analysis beyond speech. *PLoS ONE*, page in press, 2011.
- [2] M. A. Montemurro and D. H. Zanette. *PLoS ONE*, 6:e19875, 2011.
- [3] M. Montemurro and P. A. Pury. Long-range fractal correlations in literary corpora. *Fractals*, 10:451–461, 2002.
- [4] W. Ebeling, T. Pöschel, and K.-F. Albrecht. Entropy, transinformation and word distribution of information-carrying sequences. *International Journal of Bifurcation and Chaos*, 5:51–61, 1995.
- [5] W. Li and K. Kaneko. Long-range correlation and partial $1/f^\alpha$ spectrum in a noncoding DNA sequence. *Europhysics Letters*, 17(7):655–660, January 1992.
- [6] Z.-Y. Su and T. Wu. Music walk, fractal geometry in music. *Physica A: Statistical Mechanics and its Applications*, 380(0):418 – 428, 2007.
- [7] R. Ferrer i Cancho and D. Lusseau. Long-term correlations in the surface behavior of dolphins. *Europhysics Letters*, 74(6):1095–1101, 2006.
- [8] B. McCowan, S. F. Hanser, and L. R. Doyle. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Anim. Behav.*, 57:409–419, 1999.
- [9] S. A. Altmann. Sociobiology of rhesus monkeys. II: Stochastics of social communication. *J. Theor. Biol.*, 8:490–522, 1965.
- [10] G. K. Zipf. *Human behaviour and the principle of least effort*. Addison-Wesley, Cambridge (MA), USA, 1949.
- [11] R. Ferrer-i-Cancho and B. McCowan. A law of word meaning in dolphin whistle types. *Entropy*, 11(4):688–701, 2009.
- [12] R. Suzuki, P. L. Tyack, and J. Buck. The use of Zipf's law in animal communication analysis. *Anim. Behav.*, 69:9–17, 2005.
- [13] G. A. Miller and N. Chomsky. Finitary models of language users. In R. D. Luce, R. Bush, and E. Galanter, editors, *Handbook of Mathematical Psychology*, volume 2, pages 419–491. Wiley, New York, 1963.
- [14] P. Niyogi and R. C. Berwick. A note on Zipf's law, natural languages, and noncoding DNA regions. *A.I. Memo No. 1530 / C.B.C.L. Paper No. 118*, 1995.
- [15] R. Ferrer-i-Cancho and B. Elvevåg. Random texts do not exhibit the real Zipf's-law-like rank distribution. *PLoS ONE*, 5(4):e9411, 2009.

- [16] W. Ebeling and T. Pöschel. Entropy and long-range correlations in literary English. *Europhysics Letters*, 26(4):241–246, 1994.
- [17] R. Ferrer i Cancho. Euclidean distance between syntactically linked words. *Physical Review E*, 70:056135, 2004.
- [18] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley, New York, 2006. 2nd edition.
- [19] R. Durstenfeld. Algorithm 235: Random permutation. *Communications of the ACM*, 7(7):420, 1964.
- [20] J. A. Rice. *Mathematical statistics and data analysis*. Duxbury, Belmont, CA, 2007. 3rd edition.
- [21] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.
- [22] G. A. Miller. Some effects of intermittent silence. *Am. J. Psychol.*, 70:311–314, 1957.
- [23] A. Rapoport. Zipf’s law re-visited. *Quantitative Linguistics*, 16:1–28, 1982.